**Rapor: Basit Bir Tahmin Modeli Oluşturma**

**Giriş**

Bu rapor, Iris veri seti kullanılarak oluşturulan basit bir tahmin modeli projesini kapsamaktadır. Proje kapsamında veriler ön işlenmiş, lojistik regresyon modeli eğitilmiş, modelin performansı değerlendirilmiş ve sonuçlar yorumlanmıştır.

**Veri Seti**

Iris veri seti, 150 örnekten oluşur ve her örnek 4 özellik içerir: sepal length, sepal width, petal length ve petal width. Hedef değişken ise 3 farklı türden birini belirtir: setosa, versicolor veya virginica.

**Modelleme Süreci**

**1. Veri Yükleme ve Keşifsel Veri Analizi (EDA)**

İlk adımda, Iris veri seti pandas kütüphanesi kullanılarak yüklendi. Veri setinin genel istatistiksel özeti çıkarıldı ve eksik değerler kontrol edildi. Hedef değişkenin dağılımı incelendi.

**2. Veri Ön İşleme**

Veri seti, özellikler (X) ve hedef değişken (y) olarak ayrıldı. Veriler eğitim ve test setlerine ayrıldı (eğitim seti %80, test seti %20). Özellikler, StandardScaler kullanılarak standartlaştırıldı.

**3. Model Eğitimi**

Lojistik regresyon modeli sklearn kütüphanesi kullanılarak oluşturuldu ve eğitim seti ile eğitildi. Modelin eğitimi için azami iterasyon sayısı 200 olarak belirlendi.

**4. Model Performansını Değerlendirme**

Eğitim tamamlandıktan sonra, model test seti üzerinde değerlendirildi. Modelin doğruluk (accuracy), kesinlik (precision) ve duyarlılık (recall) metrikleri hesaplandı. Ayrıca, karışıklık matrisi ve sınıflandırma raporu oluşturuldu.

**Değerlendirme Sonuçları**

Modelin test seti üzerindeki performansı aşağıdaki gibidir:

* **Doğruluk (Accuracy):** 0.93
* **Kesinlik (Precision):** 0.94
* **Duyarlılık (Recall):** 0.93

**Karışıklık Matrisi**

lua

Kodu kopyala

[[10 0 0]

[ 0 8 2]

[ 0 0 10]]

**Sınıflandırma Raporu**

markdown

Kodu kopyala

precision recall f1-score support

0 1.00 1.00 1.00 10

1 1.00 0.80 0.89 10

2 0.83 1.00 0.91 10

accuracy 0.93 30

macro avg 0.94 0.93 0.93 30

weighted avg 0.94 0.93 0.93 30

**Sonuçlar ve Yorumlar**

Model, genel olarak yüksek doğruluk, kesinlik ve duyarlılık oranlarına sahiptir. Ancak, sınıf 1 için duyarlılık oranı (0.80) diğer sınıflara göre daha düşüktür. Bu, modelin sınıf 1'i diğer sınıflara kıyasla daha az doğru bir şekilde tahmin ettiğini göstermektedir.

**Potansiyel İyileştirmeler**

1. **Daha Fazla Veri:** Modelin performansını artırmak için daha fazla veri toplanabilir. Daha büyük veri setleri, modelin genelleme yeteneğini artırabilir.
2. **Diğer Modellerin Denenmesi:** Farklı makine öğrenimi modelleri (örneğin, destek vektör makineleri, karar ağaçları veya rastgele ormanlar) denenebilir ve performans karşılaştırması yapılabilir.
3. **Özellik Mühendisliği:** Yeni özellikler türetmek veya mevcut özellikleri iyileştirmek, modelin performansını artırabilir.
4. **Modelin Parametre Ayarı:** Modelin hiperparametreleri (örneğin, lojistik regresyon için düzenlileştirme parametresi) optimize edilebilir.
5. **Veri Dengesizliği İle Başa Çıkma:** Veri dengesizliği durumunda, SMOTE gibi tekniklerle veri çoğaltma veya sınıf ağırlıklarını ayarlama stratejileri kullanılabilir.

**Sonuç**

Bu proje, Iris veri seti kullanılarak basit bir tahmin modeli oluşturmanın temel adımlarını kapsamaktadır. Lojistik regresyon modeli, veri seti üzerinde yüksek performans göstermiştir. Ancak, performansın daha da artırılabilmesi için çeşitli iyileştirme yöntemleri uygulanabilir.